Constructing Semantic Campus for Academic Collaboration

Natenapa Sriharee[1] and Ravikarn Punnarut[2]

Department of Computer and Information Science
King Mongkut's Institute of Technology North Bangkok
Piboolsongkram, Bangsue, Bangkok, Thailand 10800
nsr@kmitnb.ac.th[1], rvk@student.kmitnb.ac.th[2]

Abstract: This paper proposes a methodology for constructing Semantic Campus, a Semantic Web application that represents the social network of the academics in the university, King Mongkut's Institute of Technology North Bangkok. Semantic Campus is constructed by extracting data that is available on the web site of the university. The extracted data is analyzed with respect to its association to terms defined in an ontology and associations between people in the university to reveal whether or not one academic knows another. We also discuss relation analysis that considers direct and indirect association of the campus-based resources (i.e. knowledge about the people) contained in Semantic Campus. Such analysis can be used, for example, to find specific experts in the university and research interests shared by a number of academics.

Keywords: Social Network, Academic Collaboration, Semantic Web, Semantic Campus.

1 Introduction

In sociology, using a computer network to connect people or organizations is seen as contributing to the creation of a social network, i.e. a set of social relationships in the community [1]. The social network represents a common interest of the people in the community and it can be defined by the pattern of relations between people or organizations. The relations can be described in terms of how one network member relates to another, such as being a friend, a co-worker, etc. Such descriptions make possible an analysis of collaboration, such as finding a group of people who have the ability to co-operate in doing research together by considering the research topics in which they are interested.

Semantic Web [2] is a challenging technology that can be applied as a driven mechanism to represent a social network. The description in the world of Semantic Web is explicitly described with a given well-defined meaning and as such is available for analysis and for machine query. With such a semantic description, social network analysis can be applied broadly with consideration of the knowledge contained in such a network.

This paper proposes a framework for constructing Semantic Campus, which is a Semantic Web application that represents the social network of academics in a university (in this case, our university). Academic and organizational information that is available on the web site of the university is investigated in order to create a campus-based resource described in terms of semantic description using FOAF [3]. The campus-based resource is also enriched with additional semantics that link one resource to another. Through enrichment of the social network of Semantic Campus, we aim to provide a range of capabilities such as:

- *To be able to diagnose relationships between the academics in the university.*
- *To find potential experts in specific research areas.*
- *To provide useful information that represents the individual experience of the academics and research interests that they share.*
- *To discover internal connections that can help the academics perform their activities (e.g. research, thesis committee formation).*

This paper is arranged as follows: Section 2 reviews related work, while Section 3 presents an overview of the multi-step process of constructing Semantic Campus. A detailed description of the multi-step process of constructing Semantic Campus is contained in Sections 4, 5 and 6. Section 7 discusses our approach to creating the detailed association of personal description. Section 8 introduces relation analysis and Section 9 presents conclusions and a discussion of future work.

2 Related Work

Applications relating to social networks cover broad research areas, hence we focus here specifically on related works that follow the Semantic Web approach. There are some works that consider FOAF metadata. [4] proposes an enhanced FOAF to describe more campus-based resources, [5] discovers semantics about the personal description gathered from the web and [6] seeks to ascertain the social interactions among the people on the site (e.g. blogging, LiveJournal) for visualizing FOAF. [7] and [8] use the enriched FOAF and deal with relation analysis. The former diagnoses relations to detect conflicts of interest in reviewing conference papers and the latter uses RuleML for relation analysis in an expert-finding system. Our work follows the ideas of the works mentioned above, though we also enhance the methodologies with regard to obtaining data, association of the data and ontology, creating a semantic description and analyzing it.

3 Multi-step Process of Constructing Semantic Campus

The development of a Semantic Web application typically involves several tasks such as obtaining the data, removing from it ambiguous data, enriching the data to create semantics-based data, representing it with expressive language, querying and perhaps providing a visualization, and analyzing the data.

In this paper, we define our multi-step process (see Fig. 1) of constructing Semantic Campus as follows.

(i) *Classifying sources of information for gathering the data.* In this step, we determine sources of information that will be exploited by investigating the types of web page contained in the web site of the university. We classify a web page into different domains. A domain is a source of information for gathering the data.

(ii) *Defining metadata and ontology to represent campus-based resources.* In this step, we construct an ontology for use in describing campus-based resource description by considering the data that is available on the web site of the university and enriching it with analysis of the conceptualizations it incorporates. Some automated engines can be integrated in this step.

(iii) *Creating campus-based resource description.* In this step, we extract the data from the gathered documents that are crawled. The extracted data will be mapped into terms defined in the ontology. The domain expert will define appropriate mapping, though a semi-automatic mapping tool is also provided.

(iv) *Analyzing relations defined in the social network of Semantic Campus*. In this step, an algorithm of relation analysis is defined to analyze the knowledge (i.e. facts, including new facts) contained in the social network represented by Semantic Campus.

(v) *Querying information of Semantic Campus.* In this step, a browsing approach will be developed in which the user can define the concepts and properties for querying against the knowledge defined in Semantic Campus. Visualization can be provided by linking information from Semantic Campus to information in the current web site.

Even though the functionality of each step is distinctly defined, the multi-step process should have the ability to re-enact the whole process or individual steps when the description (e.g. data on the web) changes.

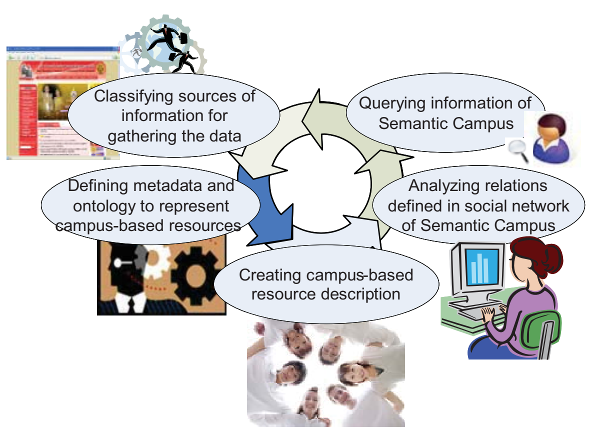

Fig. 1. Multi-step process of application

4 Classifying Sources of Information for Gathering Data

In this research, we collect the data by using the web map of the *www.kmitnb.ac.th* web site. The web pages contained in the web site are mainly described in Thai language. We use a link extractor to extract links from the relevant web pages defined in the web map and such links (i.e. initial documents) are classified into different type of web pages. A Campus Site Structure tree is created, which contains the links (represented as nodes) that may entail a description of campus-based resources (e.g. personal description, organizations) and will be used for gathering the data. The web pages themselves are further investigated at a deeper level and additional links will then be added into the Campus Site Structure tree. Since the web pages are implemented with heterogeneous web application technology and have different structures, we provide a parser to verify whether such web pages may refer to more information defined in another web page. The construction of Campus Site Structure tree can be considered with regard to the structure of university organizations: for example, a faculty web page can be defined as an upper level node while its department web pages can be defined as the lower level nodes. A node is defined in terms of its depth and its parent node. For example, personal web pages can be accessed through faculty, departmental and personal web pages, with the depths being specified as 1, 2 and 3, respectively.

Through prior investigation we discovered how frequently a person description appears in a web page by searching person names using prefixes (e.g. ‘Professor’, ‘Mr.’, ‘Ms.’) that identify the persons. Distribution of the discovered person names from 760 links with respect to different type of web page and depth (the number of web pages being specified in parentheses) is contained in Table 1.

Table 1. Distribution of person names in the web pages of university web site

Node	Initial documents	Depth 1	Depth 2	Depth 3	Depth 4	Depth 5
Root (depth 0) (*www.kmitnb.ac.th*)	1 (1)	-	-	-	-	-
Faculty	49	0 (1)	0 (9)	896 (95)	1,633 (314)	2,976 (198)
Officer	24	0 (1)	0 (12)	87 (29)	62 (13)	219 (20)
Chief Executive Officers	8	0 (1)	46 (7)	-	-	-
Research Center	8	0 (1)	0 (7)	4 (3)	-	-
Student Information Service	12	0 (1)	0 (11)	-	-	-
University Web Service	16	0 (1)	0 (15)	-	-	-
Admission Information	3	0 (1)	0 (2)	-	-	-
About	12	9 (5)	0 (7)	-	-	-
Other Web Pages	5	3 (5)	-	-	-	-
Total	138					

5 Defining Ontology for Describing Personal Description

An ontology can be constructed by gathering terms in existing systems (e.g. [9] and [10]) and by the domain expert who is responsible for describing the ontology. We follow an approach of constructing an ontology by investigation of terms that are available on the web and define it by considering IS-A relationship between concepts that should be defined in the university domain. This paper uses FOAF [3] in describing instances of personal description. FOAF defines simple metadata for describing associations between people and organizations, such as *foaf:knows*, *foaf:currentProject*, *foaf:fundedBy*, etc. In order to come up with a useful initial set of terms, the target audience should be identified and a set of use cases should be defined. The use cases relate to which information from the Semantic Campus should be provided to meet users' requests. For example, use cases could be: finding an academic who is an expert in a specific research area and finding the academics or researchers who have a common interest in research topics relating to Semantic Web.

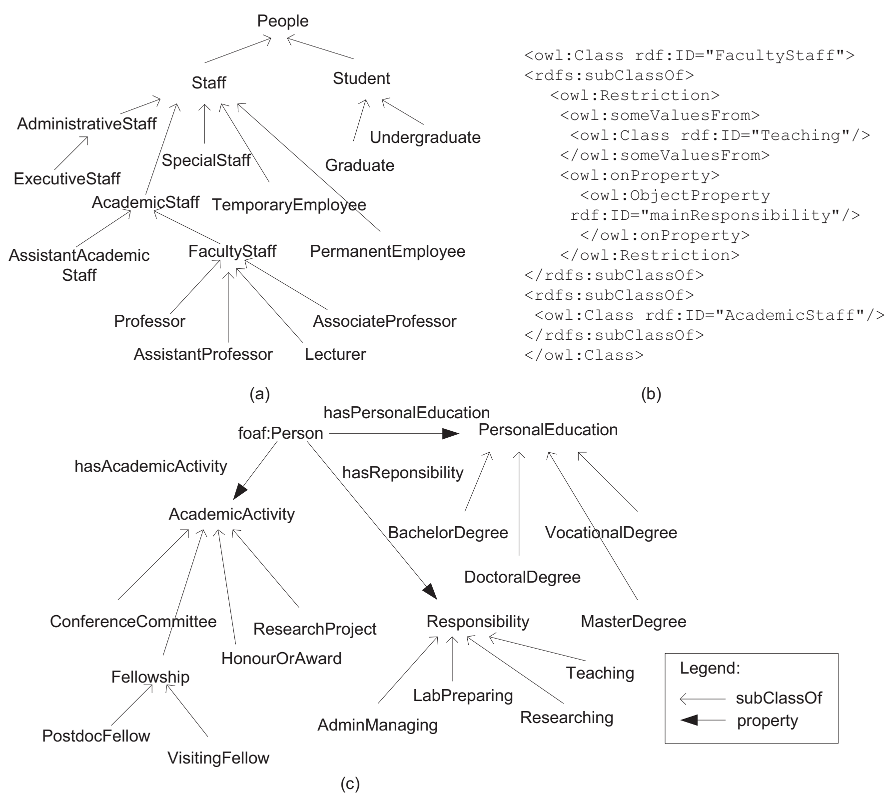

Fig. 2. (a) Partial terms in ontology (b) Person's role describing (c) Person's detailed view

Currently, the university has six types of staff: academic staff, assistant academic staff, administrative staff, special staff, permanent employees and temporary employees [11]. Each type has a different role and responsibility, and different conditions of employment. Fig. 2 (a) shows partial terms defined in Semantic Campus ontology, the ontology for describing personal descriptions of the people in the university. A type of staffs is represented by role. Fig. 2 (b) represents a restriction constraint that specifies that a person's role is "faculty" if his/her main responsibility is teaching.

We represent personal description at an abstract level as a graph with edges associating to a set of vertices that are partitioned into several disjoint sets corresponding to semantics defined in the ontology. For example, prior personal description can be general person description (i.e. described by core FOAF metadata), responsibility, education, academic activity, position, research topic, etc. Due to limitation of space, we simply illustrate personal description at an abstract level with three sets of vertices, i.e. academic activity, responsibility and person education, as shown in Fig. 2 (c).

6 Overview of Creating Campus-based Resource Description

Fig. 3 represents the process of creating a campus-based resource description, involving retrieving documents, extracting data, computing an association matrix and generating a personal description. Finding an extraction methodology that can deal with Thai language is a difficult task because there are no explicit word boundaries in Thai [12], therefore we follow an approach of keyword-mapping using a term set similar to [5] but applying the process to Thai terms. The term set has a set of terms that are relevant to terms defined in Semantic Campus ontology and such terms will be used to search the description of campus-based resources (e.g. persons, organizations). Mapping terms can be performed by calculating the similarity between two lexical terms. It is possible that an extracted term may match by similarity to more than one term in the ontology, in which case the domain expert will specify appropriate mapping to only one term.

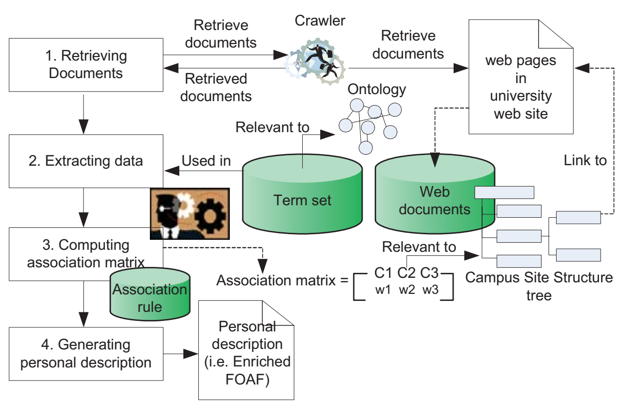

Fig. 3. Creating personal description

A detailed description of the association between extracted terms from personal web pages will be produced by determining the relevancy between the terms and such is represented by association matrix (see Section 7). The result of creating campus-based resource is semantic descriptions of the people in terms of the enriched FOAF, involving personal description links to semantics defined in the ontology as well as having detailed association.

7 Creating Detailed Association of Personal Description

We propose the use of an association matrix to display relationships between the two campus-based resources that are being considered based on their discovered context (e.g. web page). In this paper we classify the discovered context into two types (shared context and referenced context) on the basis of the following criteria:

(i) The shared context is a web page that refers to the two campus-based resources' description that is being considered. For example, person P1's web page mentions the name of P2, or the department web page mentions their names. A shared context can be a specific context of one: for example, the previous example demonstrates that P2 is known by P1 in the specific context of P1 (i.e. the personal web page of P1).

(ii) The referenced context is a web page (i.e. upper level node) that refers to another web page (i.e. lower level node) defined in the Campus Site Structure tree (mentioned in Section 4), but does not explicitly mention the two campus-based resources' description. For example, a faculty web page is the referenced context of the personal web pages of P1 and P2 but it does not mention the names of P1 and P2.

Association of the resource R_i and the resource R_j in the discovered context (C_1 *to* C_n) that is relevant to R_i can be represented by notation (1) as follows:

$$associationType\left(R_i,\ R_j\right) = \left[\alpha_{C_1},...,\alpha_{C_n}\right] \quad \textbf{(1)}$$

Where association type represents the relation between the two resources being considered and $\alpha_{C_1},...,\alpha_{C_n}$ represent the weight of the relevancy of the resources in a discovered context. The weight of discovered context is specified regarding the depth and structure of links defined in the Campus Site Structure tree and is stored in an association rule database (Fig. 3).

Fig. 4 (a) represents 'knows' as association type. In the discovered contexts that are relevant to P1, for example, such associations can be represented by $knows(P1,P2) = [0.5_{fac}\ \ 0.8_{dep.}\ \ 1_{P1web}]$ where 0.5, 0.8 and 1 are weights for their descriptions that are discovered in the referenced context (i.e. faculty web page) and the shared contexts (i.e. department web page and P1 web page), respectively. Therefore, the confidence level that determines 'P1 knows P2' will be the maximum weight of these discovered contexts (in this case, 1), while 'P2 knows P1' has a different confidence level of 0.8. Specifying confidence of association can be represented by notation (2) as follows.

$$confidenceOfAssociation(R_i, R_j) = Max(\alpha_{C_1}, ..., \alpha_{C_n}) \quad \textbf{(2)}$$

Direction of association between two resources can be a uni-directional relation or bi-directional relation. In this paper we determine the direction of relation by considering the discovered context with conditions as follows:

(i) If the resources are discovered in a shared context but that context is not specific to them, it means that their relation can be bi-directional (e.g. both P1 and P2 are mentioned in a list of staffs in the same department web page).

(ii) If the resources are discovered in the specific context of one resource, it means their relation can be uni-directional (e.g. the personal web page of P1 mentions P2).

(iii) If the resources are discovered in the referenced context it means their relation can be bi-directional (e.g. the faculty web page refers to personal web pages P1 and P2 in the Campus Site Structure tree)

It is possible that there is more than one satisfied condition. For example in Fig. 4 (a), conditions (i), (ii), and (iii) are satisfied in the relevant contexts of P1, hence we give a higher priority to bi-directional relation rather than uni-directional relation. In this case, therefore, the relation between P1 and P2 is constituted by bi-directional relations with maximum weight as 1 to represent a confidence or probability of knowing each other.

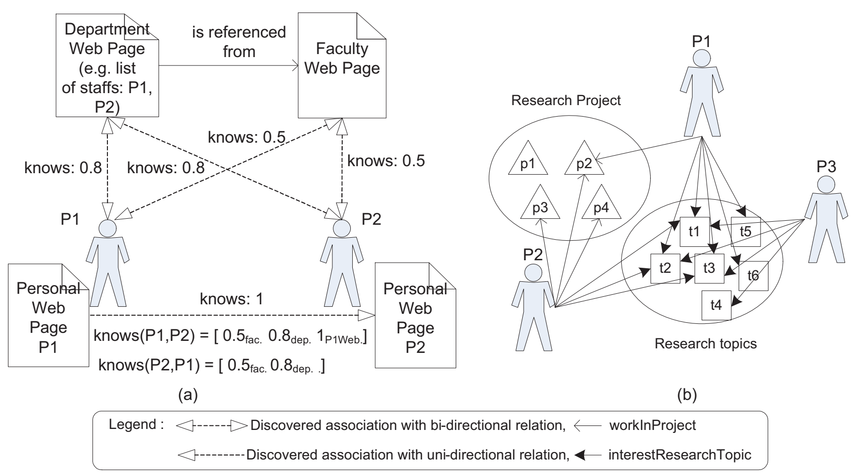

Fig. 4. (a) Association between people (b) Example of social network

8 Relation Analysis

Relation analysis is a methodology to discover relevancy between the resources by considering direct relations and indirect relations. In this work we conduct direct relation analysis through querying technique and indirect relation analysis through association reasoning. An example of the former is represented in Fig. 4 (b), and involves finding the academics who have a common interest in the research topics t1, t2 and t3 to form a group of researchers for a research project (that includes academics P1, P2 and P3). We can do this straightforwardly through querying the semantic relation (i.e. interestResearchTopic) of each person. To consider how close the retrieved information is to the query can be considered by weight of association ranking (e.g. [13]). Relation analysis is useful in various ways if allowing people to request connection to others; for example, if P3 wants to make connections with P1 and P2, he/she can request a connection to both (e.g. www.linkedin.com) and once their links are established they become part of the knowledge embedded in the social network, enabling various methods of querying. We give the assertions defined in the Semantic Campus as follows.

- *knows (P1,P2)*
- *interestResearchTopic (P1, 'Semantic Web')*
- *interestResearchTopic (P2, 'Semantic Web')*
- *workInProject (P1, ProjectA)*
- *includeResearchTopic ('ProjectA, 'Semantic Web')*

The association reasoning can be conducted by considering these assertions. For example finding out if P2 is a co-worker of P1 in project A can be achieved by considering relevant knowledge to see if it provides sufficient support for the hypothesis that P2 may be a co-worker of P1. Belief network [14] can be incorporated into association reasoning. With belief network, relation analysis can consider a new association between the two resources using the existing assertions, and this can provide the confidence of the newly discovered assertion. In this example, it is possible to decide that P2 is a co-worker of P1 in project A if the weight of their assertions (i.e. P2 interests in Semantic Web and the inclusion of Semantic Web as a research topic in project A) strongly confirms the belief.

9 Conclusion and Future Work

In this paper, we describe our framework for constructing Semantic Campus that will be used for providing information to the academics and students in the university. Semantic Campus is created using data collected from existing systems. The multi-step process of constructing Semantic Campus is proposed. Attention has been paid to identifying sources of information for gathering the data, so that the obtained data can represent meaningful semantics and hence is beneficial for analysis.

Many crucial tasks in developing Semantic Campus remain to be completed, such as an evaluation of the capability of a semi-automatic tool that aids in mapping terms of the extracted data to terms defined in the ontology and creating semantic information. However, the extracted data often lacks detailed description and hence may not be sufficient for discovering associations and conducting relation analysis. Hence, we aim to integrate other sources of information with Semantic Campus, such as the research reports of the National Research Council of Thailand and academic publications stored in the research database of the university. Relation analysis in our future work will be conducted by following the belief network approach.

Acknowledgements. The authors would like to acknowledge King Mongkut's Institute of Technology North Bangkok for support of this work.

References

1. Garton, L., Haythornthwaite, C., Wellman, B.: Studying Online Social Networks. Journal of Computer Mediated Communication, 3 (1) (1997)
2. Semantic Web, http://www.w3.org/2001/sw/
3. Brickley, D., Miller, L.: FOAF Vocabulary Specification 0.9, May 24, 2007, http://xmlns.com/foaf/0.1/
4. Nowack, B.: Semantic Campus - A FOAF Extension. Position paper for FOAF 2004 in Galway, Ireland, September (2004)
5. Mori, J., Matsuo, Y., Ishizuka, M., Faltings, B.: Keyword Extraction from the Web for FOAF Metadata. In: Proceedings of the 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web, Galway, Ireland (2004).
6. Paolillo, John C. and Wright, E.: Social Network Analysis on the Semantic Web: Techniques and Challenges for Visualizing FOAF, EServer TC Library, http://tc.eserver.org/25494.html, 2004
7. Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A.P., Arpinar, I.B., Joshi, A., Finin, T.: Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection. In: Proceedings of the 15th International World Wide Web Conference (WWW2006), Edinburgh, Scotland, UK, May (2006)
8. Li, J., Boley, H., Bhavsar, V.C., and Mei, J.: Expert Finding for eCollaboration Using FOAF with RuleML Rules. In: Proceedings of Montreal Conference on eTechnologies (MCTECH), Canada. pp. 53-65 (2006).
9. Mika, P.: Ontologies are us: A Unified Model of Social Networks and Semantics. In: 4th International Semantic Web Conference (ISWC 2005), Galway, Ireland, November, pp. 522–536 (2005).
10. Patel, C., Supekar, K., and Lee, Y.: OntoGenie: Extracting Ontology Instances from WWW. In: Workshop at ISWC 2003 International Semantic Web Conference Sanibel Island, Florida, October (2003)
11. Information of Academic, http://www.kmitnb.ac.th/kmitnb_it/
12. Sornlertlamvanich, V., Potipiti, T. and Charoenporn, T.: Automatic Corpus-Based Thai Word Extraction with the C4.5 Learning Algorithm. In: 18th International Conference on Computational Linguistics COLING (2000), Saarbrucken, Germany, pp 802-807, July-August (2000).
13. Aleman-Meza, B., Halaschek-Wiener, C., Arpinar, I.B., Ramakrishnan, C., and Sheth, A.P.: Ranking Complex Relationships on the Semantic Web. IEEE Internet Computing, 9(3):37-44, May/June (2005)
14. Neapolitan, R.E.: Learning Bayesian Networks. Prentice Hall, Upper Saddle River, NJ (2003)